Sādhanā (2019) 44:54
https://doi.org/10.1007/s12046-019-1052-x

Efficient band selection for improving the robustness of the EMD-based cepstral features

EHSAN SAMADI and GHASEM ALIPOOR*

Electrical Engineering Department, Hamedan University of Technology, Hamedan 6516913733, Iran
e-mail: samadi@stu.hut.ac.ir; alipoor@hut.ac.ir

MS received 13 April 2018; revised 1 December 2018; accepted 3 December 2018; published online 9 February 2019

Abstract. Mel-Frequency Cepstral Coefficients (MFCC) are features widely and successfully used for various speech processing applications. These features are extracted using Fourier transform. However, this transform suffers from some crucial restrictions when used for analyzing nonlinear and non-stationary signals such as speech. To address this problem, in the present study, we investigate the application of Empirical Mode Decomposition (EMD) in extracting more efficient and robust features for automatic gender identification. In particular, in the proposed approach, the speech signal is first decomposed into a set of narrow-band oscillatory modes, using EMD, from which mel-frequency cepstral features can be extracted. On the other hand, multi-band decomposition of all modes results in some redundant and even irrelevant features that can degrade the performance of the classification. Therefore, we propose to efficiently select the most discriminative frequency bands over all modes. The minimal-redundancy-maximal-relevance (mRMR) feature selection algorithm is also examined for this purpose. The proposed EMD-based features are then extracted by applying DCT on log power values calculated over the selected mel-scale bands of the IMFs. Simulation results show that, using the proposed features for automatic gender identification considerably improves the performance of the system, in particular in noisy environments.

Keywords. Automatic gender identification; empirical mode decomposition (EMD); mel-frequency cepstral coefficients (MFCC); feature selection; minimal-redundancy-maximal-relevance.

1. Introduction

The aim of a speaker gender identification system is to automatically identify the gender of the speaker. Such a system has numerous practical applications. For instance, it can be helpful in sorting out the incoming phone calls on the basis of the speaker's gender in order to provide gender-oriented services. Furthermore, as a pre-processing unit, gender identification can enhance the accuracy of some recognition models, e.g., within the speaker identification [1], speaker verification [2] and speaker diarization [3] systems, by reducing the search space.

Extraction and selection of the best parametric representation of acoustic signals is an important task in the design of any gender identification system. Pitch [4–6] and Formant [7] frequencies are two speech-specific features widely used for this purpose. Some other features that have been studied in this field are correlation coefficients, Fourier-Bessel coefficients [8] and mel-frequency cepstral coefficients (MFCCs) [9–11]. Furthermore, various classification methods have been employed for this purpose, e.g., artificial neural networks (ANN) [6], Gaussian mixture models (GMM) [4, 9, 12], linear discriminant analysis (LDA) [11] and support vector machines (SVM) [8, 10]. These studies are summarized in table 1 in which the adopted features, models and datasets are included along with the achieved results. These studies have been mostly conducted for clean speech signals and mostly result in accuracy rates of about 95%.

Cepstral coefficients are among the features vastly and successfully used for speech and speaker recognition. In particular, MFCCs are short-time features extracted by applying the cosine transform on the log power spectra estimated over mel-scale-based bands. However, these features are extracted using Fourier transform that suffers from some crucial restrictions for analyzing nonlinear and non-stationary signals such as speech. To address this problem, in the present study, we investigate the application of empirical mode decomposition (EMD) in extracting more efficient and robust features for automatic gender identification. *EMD is a* robust adaptive *time-frequency analysis method* for representing a non-stationary signal as sum of a set of narrow-band oscillatory modes. These modes are called intrinsic mode functions (IMF), each with slowly varying amplitude and phase [13]. IMFs are generally in good agreement with intuitive and physical signal

*For correspondence

Table 1. A brief summary of the studies reported on gender identification.

Ref.	Features	Classifier and model	Accuracy and achievement	Dataset
[7]	Formant and fundamental frequencies	ANOVA	Formant Freq.: 98.1% Fund. Freq.: 96.2%	Personally recorded data
[6]	Acoustic and pitch features	Neural networks	93%	Recordings from French and English radio stations
[4]	Pitch and RASTA-PLP	GMM	98%	TIMIT and some other multilingual speech samples
[8]	Fourier–Bessel coefficients	SVM	About 72.92%	Personally collected data
[9]	MFCC	GMM	96.4%	TIMIT
[12]	Tone and energy variations	GMM	98.9%	Lwazi
[11]	MFCC-shifted delta coefficients	PLDA	97.63%	Fisher English (FE) and DARPA RATS corpora
[10]	MFCC & Delta MFCC	GMM-SVM	79.18%	OGI Kids
[5]	Pitch and spectral features	Logistic regression	93.8%	HMIHY

interpretations. Moreover, EMD has proven to be quite versatile in a broad range of applications for extracting information from data generated in noisy nonlinear and non-stationary processes. Furthermore, it has been shown that the EMD essentially acts as a dyadic filter bank resembling those involved in wavelet decompositions [14].

In the proposed approach, the speech signal is first decomposed into its IMFs, from which mel-frequency cepstral features can be extracted. On the other hand, multi-band decomposition of all modes results in some redundant and even irrelevant features, which can degrade the performance of the classification. Therefore, we propose to efficiently select the most discriminative frequency bands over all modes. The minimal-redundancy-maximal-relevance (mRMR) feature selection algorithm is also examined for this purpose [15]. The proposed EMD-based features are then extracted by applying DCT on log power values calculated over the selected mel-scale bands of the IMFs. The performance of the proposed EMD-based cepstral features for automatic gender identification is compared with that of the original MFCC features, in noise-free as well as noisy conditions.

This paper is presented in the following order. We give a brief description of the EMD and its variants in sections 2. The proposed EMD-based cepstral features are developed in section 3. Section 4 is dedicated to the simulation results. Conclusion remarks are presented in section 5.

2. Empirical mode decomposition (EMD) and its variants

EMD is a robust spectral *decomposition method first developed by* Huang *et al* to adaptively decompose non-stationary signals into their intrinsic oscillatory components, i.e., IMFs [13]. The EMD generalizes the Fourier analysis. Sinusoidal basis functions used in the Fourier analysis are generalized to data-dependent IMFs. Compared to a sinusoidal function, an IMF satisfies the generalized alternating property and the generalized zero-mean property, and relaxes the amplitude and frequency from being constant to being generally time-varying. Since introduction in its original form in 1998, EMD has received some evolutions. Ensemble EMD (EEMD) was introduced as a solution to the mode mixing problem that EMD frequently suffers from [16]. Another variant is the complementary EEMD (CEEMD), which possesses the completeness property [17].

2.1 *Empirical mode decomposition*

Empirical mode decomposition is based on a simple premise that each signal is made of a number of intrinsic mode functions each with the following two conditions:

- Alternating Property: The number of extrema and the number of zero-crossings must be either equal or different at most by one, i.e., IMFs have alternating stationary points and zeroes.
- Zero-Mean Property: At any point, the mean value of the envelopes defined by the local maxima and local minima is zero, i.e., the maxima and the minima of the IMFs are opposite in sign.

These IMFs are extracted through an algorithm known as the sifting process, which can be summarized as follows:

(1). Identify the local maxima (minima) of the signal, and then form the upper (lower) envelope by connecting all maxima (minima) by a curve, usually a cubic spline curve.

(2). Form the mean envelope $m_1(t)$ by averaging these upper and lower envelopes.
(3). Subtract the mean envelope $m_1(t)$ form the signal to from the first probable component as:

$$h_1(t) = x(t) - m_1(t) \tag{1}$$

(4). Ideally, $h_1(t)$ should be an IMF. But, if $h_1(t)$ does not satisfy the above definition of an IMF, let $h_1(t)$ be the new signal and repeat the steps 1-3 until the first IMF is extracted. Call the first IMF $c_1(t)$.
(5). Let $r_1(t) = x(t) - c_1(t)$. Treat $r_1(t)$ as a new signal and repeat the steps 1-4 to extract other IMFs.
(6). Repeat this procedure K times, until $r_K(t)$ is smaller than a predetermined threshold or becomes a monotonic function that no more IMF can be extracted from.

Lastly, the signal $x(t)$ is decomposed into K IMFs $c_1(t)\ldots c_K(t)$ and a residue $r_K(t)$, which can be either the mean trend of the signal or a constant. The signal $x(t)$ is composed by summing up all these components as:

$$x(t) = \sum_{k=1}^{K} c_k(t) + r_K(t) \tag{2}$$

This procedure can effectively sift the complex signals in time domain. IMF components provide valuable information about the signal. Figure 1 shows two examples of decomposing a frame of speech signal into its first 5 IMFs and its residual, for male and female speakers, using the EMD algorithm.

2.2 *Complementary ensemble empirical mode decomposition (CEEMD)*

Ensembled EMD was introduced in 2009 to tackle the mode mixing problem of the EMD. For more information about this problem one can refer to [16]. EEMD algorithm is simple and can be summarized in the following steps:

(1). Add a white noise to the signal.
(2). Decompose the noisy signal into its IMFs, using the EMD procedure.
(3). Repeat steps 1 and 2 with different samples of the white noise.
(4). Obtain the final EEMD-based IMFs as the ensemble means of the corresponding EMD-based IMFs in successive trials.

As a result of adding noise to the signal, signal reconstructed from the IMFs of the EEMD algorithm contains some residual noise. Furthermore, adding different samples of the white noise may lead to different IMFs. These issues are addressed in a modified version of the EEMD, called complementary EEMD (CEEMD). Now we try to describe this algorithm in brief. To distinguish between the IMFs resulted from the EMD and CEEDM procedures, the kth IMF based on these decomposition methods are indicated by $\mathrm{c}_k(t)$ and $\overline{\mathrm{c}_k(t)}$, respectively. $E_k(\cdot)$ is also defined as an operator that returns the kth IMF of a signal using the EMD algorithm. Furthermore, $\omega^m(t)$ is assumed to be the mth realization of a zero-mean unit-variance white noise and ε is a constant. Using these definitions, CEEMD can be summarized as follows [17].

(1). Extract the first IMF of the signal $x(t) + \varepsilon\omega^m(t)$ based on the EMD method, M times using different realizations of the noise ω. The first CEEMD-based IMF is calculated as:

$$\overline{c_1(t)} = \frac{1}{M}\sum_{m=1}^{M} c_1^m(t) \tag{3}$$

$c_1^m(t)$ is the first EMD-based IMF of the signal $x(t) + \varepsilon\omega^m(t)$ in the mth trial. The first residue is then calculated as:

$$r_1(t) = x(t) - \overline{c_1(t)} \tag{4}$$

(2). For k = 2, 3, …, the kth CEEMD-based IMF is calculated as the ensemble mean of the kth IMFs of the signal $r_{k-1}(t) + \varepsilon E_{k-1}(\omega^m(t))$ for M different realizations of the noise ω, i.e.,:

$$\overline{c_k(t)} = \frac{1}{M}\sum_{m=1}^{M} E_k(r_{k-1}(t) + \varepsilon E_{k-1}(\omega^m(t))) \tag{5}$$

(3). The residue is defined, at each iteration, as:

$$r_k(t) = r_{k-1}(t) - \overline{c_k(t)} \tag{6}$$

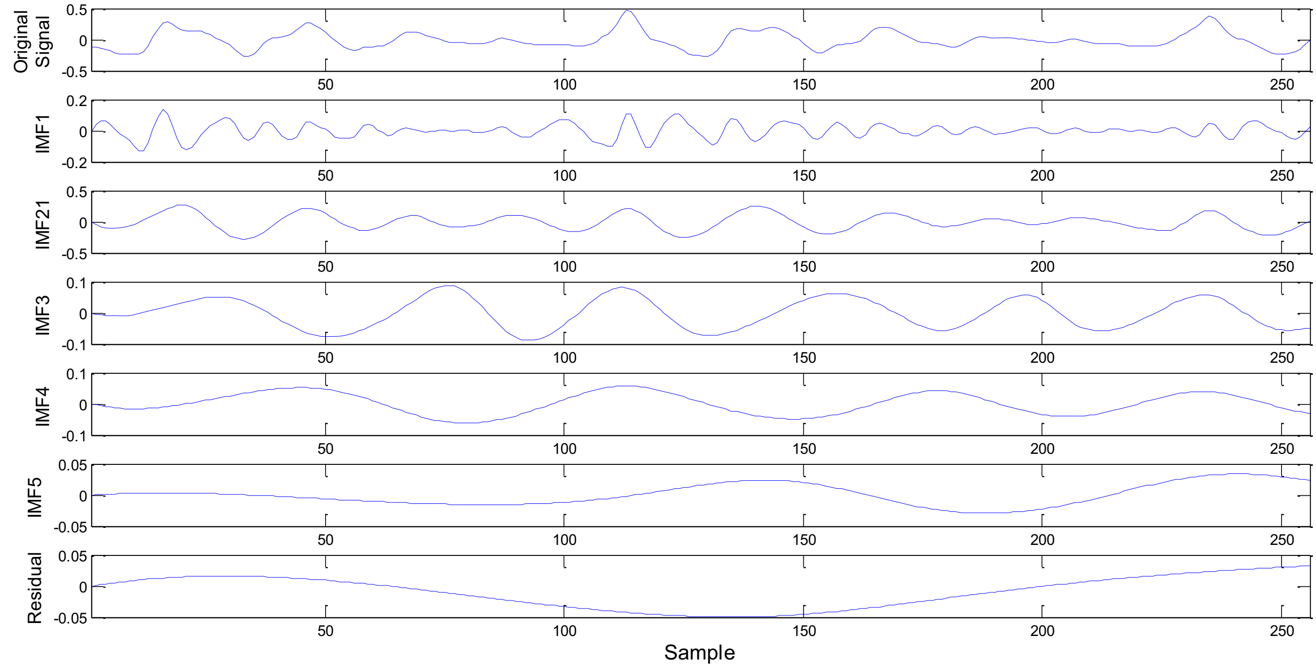
This procedure is repeated whilst the residue $r_k(t)$ has more than three extrema. The decomposed signal can now be expressed as:

$$x(t) = \sum_{1}^{k} \overline{c_k(t)} + r_K(t) \tag{7}$$

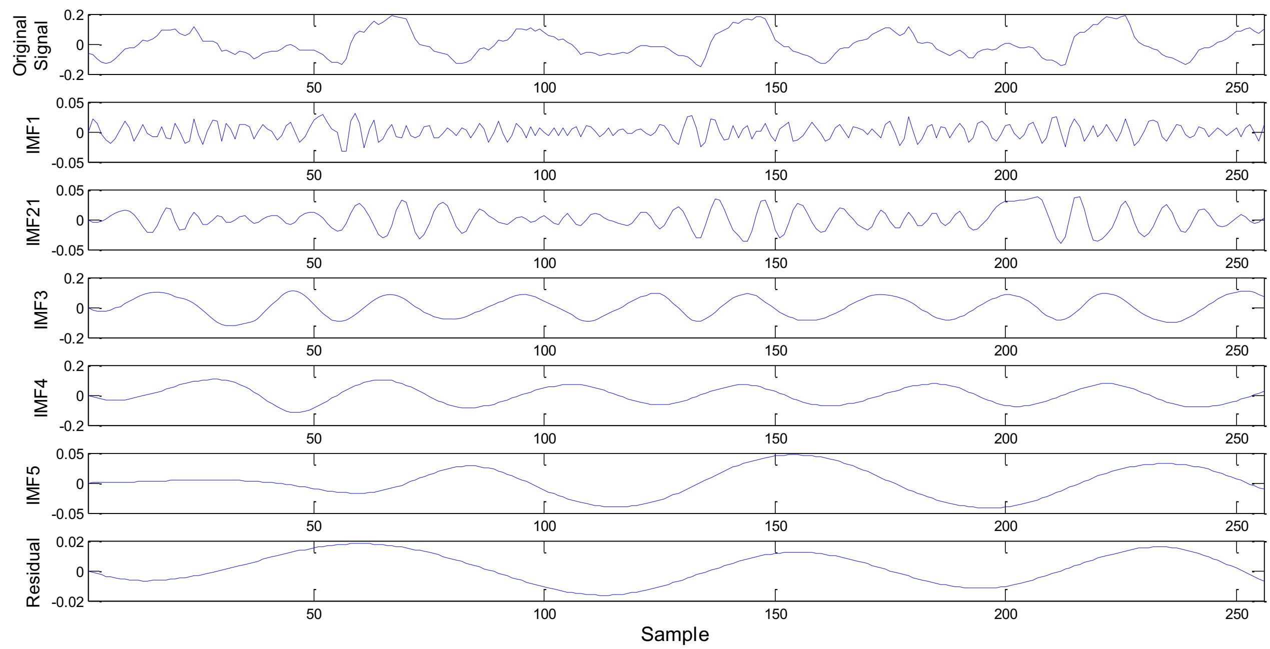
3. Cepstral features

MFCCs are short-time features extracted by applying the cosine transform on the log power spectra estimated over mel-scale-based bands. Nonlinear sub-band decomposition in accordance to the mel scale is to cope with the human auditory system. This nonlinear scale relates to the linear scale in Hertz as:

$$f_{mel} = 2595 \log_{10}\left(1 + \frac{f_{Hz}}{700}\right) \tag{8}$$



(**a**) Male speaker

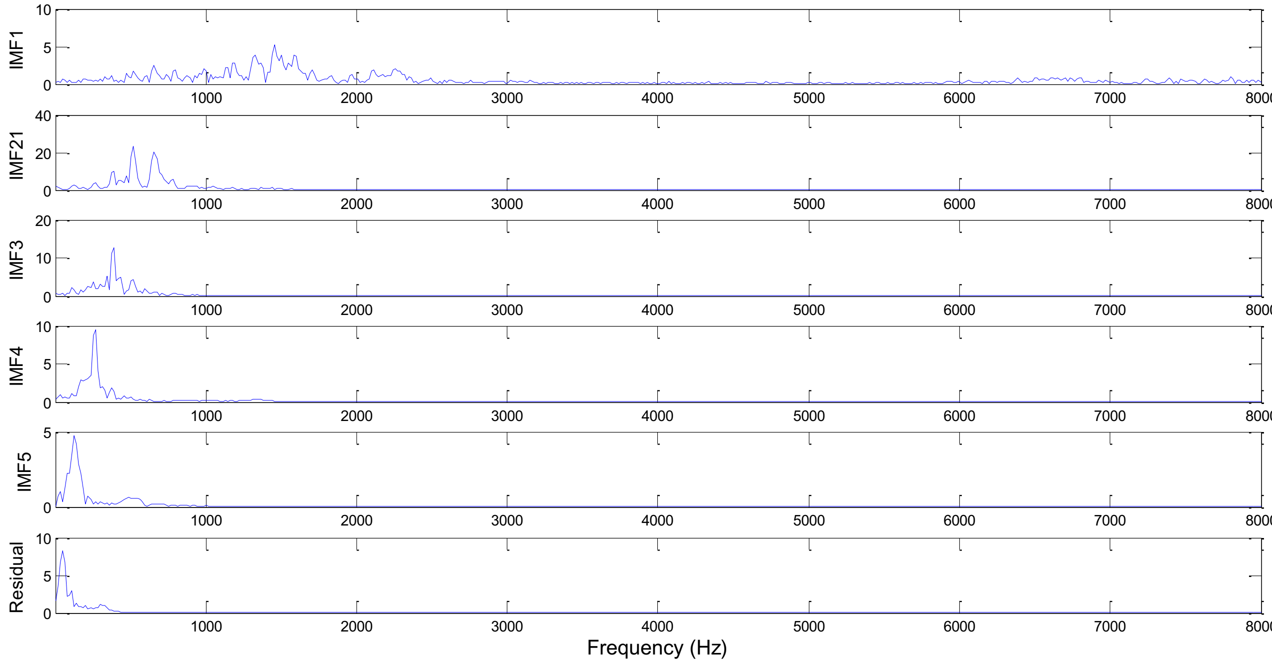


(**b**) Female speaker

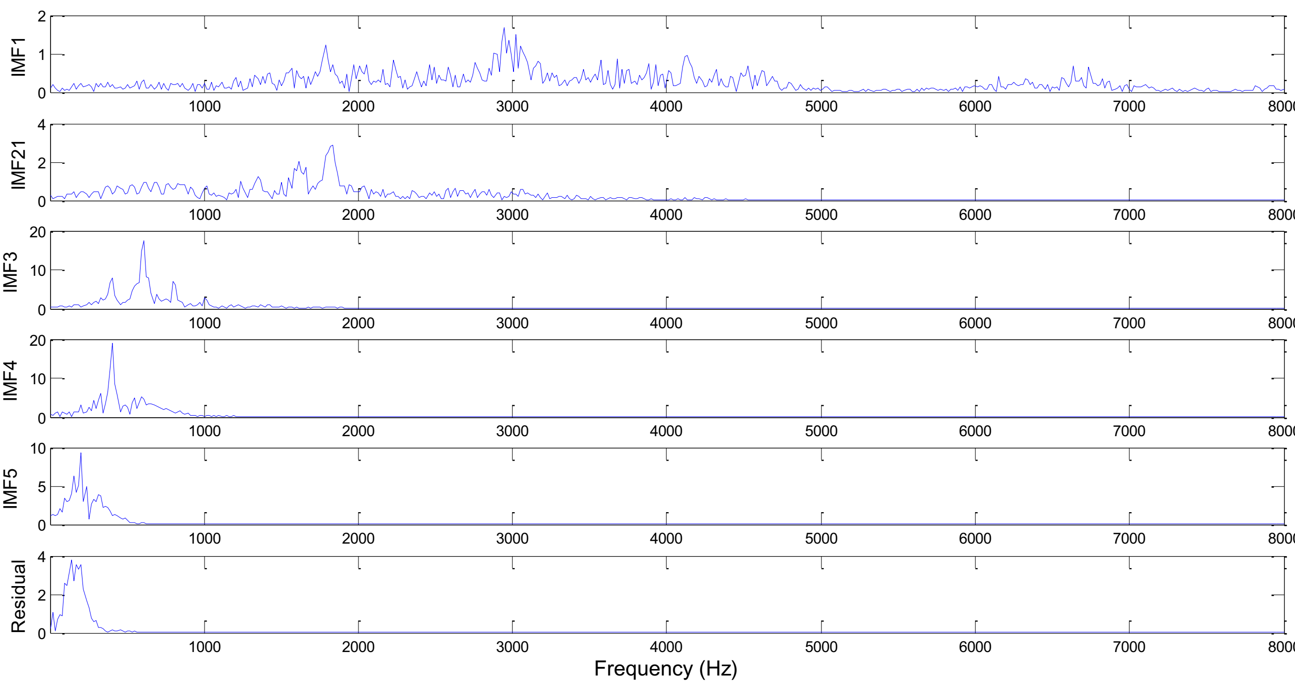
Figure 1. Decomposition of a frame of speech signal into its first 5 IMFs using the EMD, for **(a)** male and **(b)** female speakers.

Based on the commonly used procedure, MFCCs are extracted as follows.

1. The signal is first segmented into, usually overlapping, frames and each frame is then weighted by an appropriate windowing function. Over each frame, following steps are taken.
2. The spectrum of the input signal over each frame is estimated using the DFT.
3. The estimated power spectrum is mapped onto the mel scale, by applying a filter bank, with triangular overlapping frequency responses. The bandwidths as well as the central frequencies of these bandpass filters are distributed in accordance to the equation (8).



(**a**) Male speaker



(**b**) Female speaker

Figure 2. EMD-gram of the sample frames of figure 1.

4. Logarithm of the averaged power spectrum over each band is calculated.
5. Discrete cosine transform (DCT) is applied on the resultant log power spectra over all bands. The first coefficient is usually discarded and the remaining coefficients make up the MFCCs.

It is known that the MFFC features are very sensitive to additive noise. Studies have been done to alleviate this

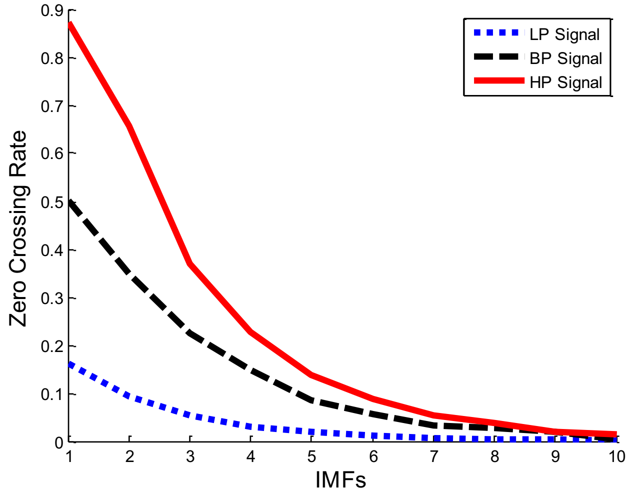


Figure 3. Zero-crossing rates of IMFs of three signals with different frequency contents.

sensitivity and improve the robustness of these features in some applications, e.g., speaker identification [18] and speech recognition [19]. In this section, using the EMD, we develop some new and improved MFCC features to address this problem.

It can be shown that EMD, as well as its variants, act as a dyadic filter bank in which the high-frequency contents of the signal reside mostly in the first IMFs and the last modes are usually of slowly-varying nature. This fact is illustrated in figure 2 in which the frequency contents of the IMFs, i.e., the EMD-gram, of the sample frames of figure 1 are depicted. Furthermore, zero-crossing rates of the extracted IMFs of three typical signals with different frequency contents are depicted in figure 3. It can be seen that going from the first IMFs to the last ones, lower-frequencies dominate and the zero-crossing rates, as an evidence of the frequency contents of the IMFs, declines. This is in turn another proof of the fact that the EMD acts as a filter bank.

Based on these observations, in this paper a modification is made on the original procedure for MFCC features extraction. In the proposed approach the speech signal is first decomposed into its IMFs. The EMD-based MFCCs are then extracted by applying DCT on log power values calculated over some specific bands of the IMFs, selected according to each band's discriminative ability. The developed algorithm can be summarized as follows:

1. Each (windowed) frame of the signal is first decomposed into 5 IMFs, using the EMD. This is because most speech signal frames can be efficiently decomposed into 5 IMFs with a negligible residue. Since common noises are usually of high frequency, it was seen that excluding the first IMF from consideration improves the

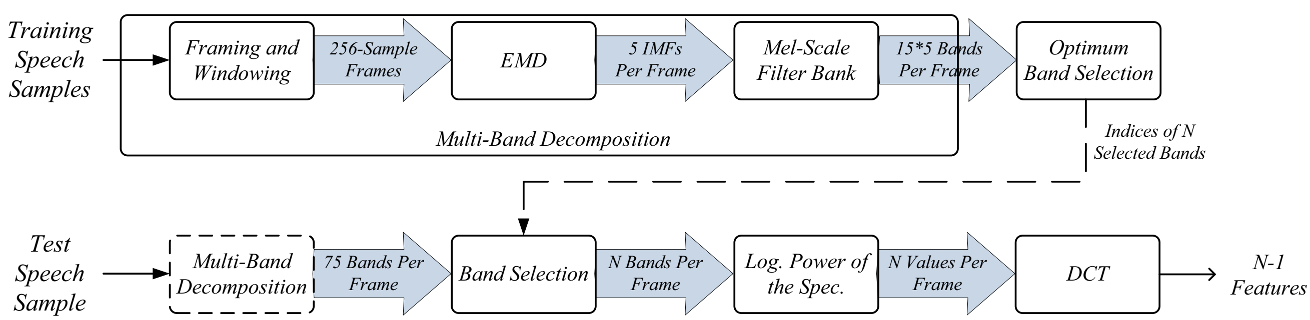
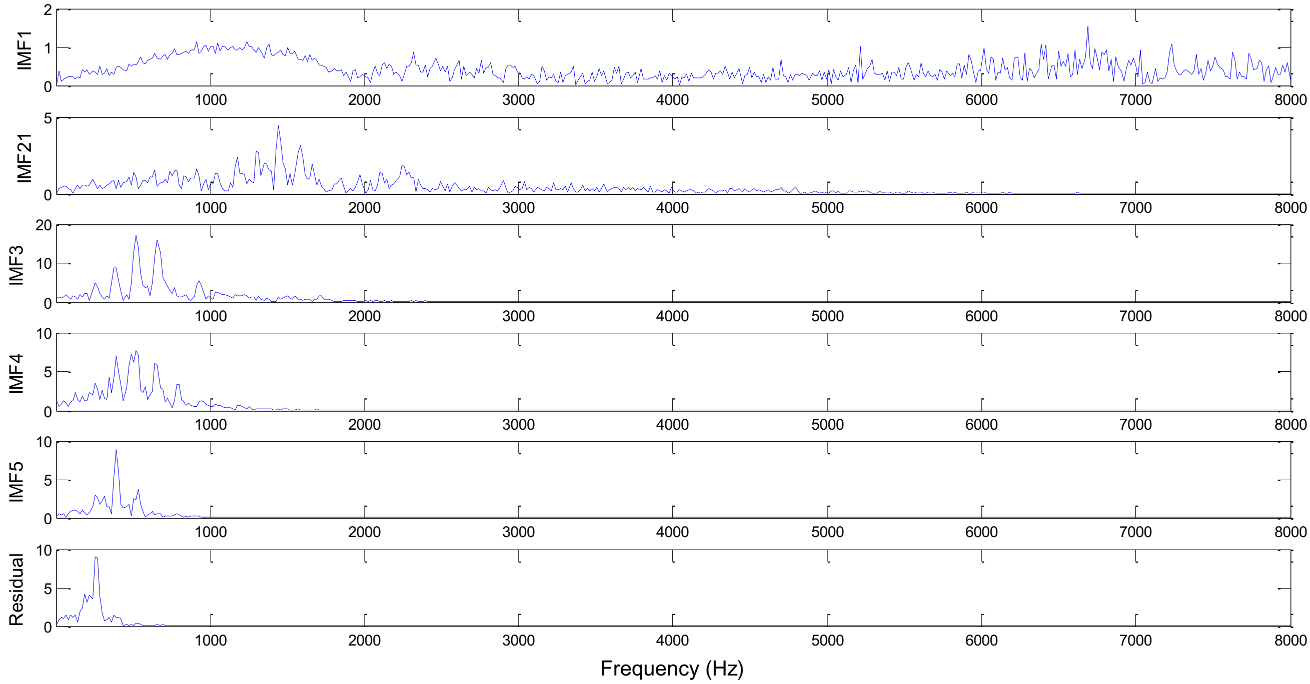


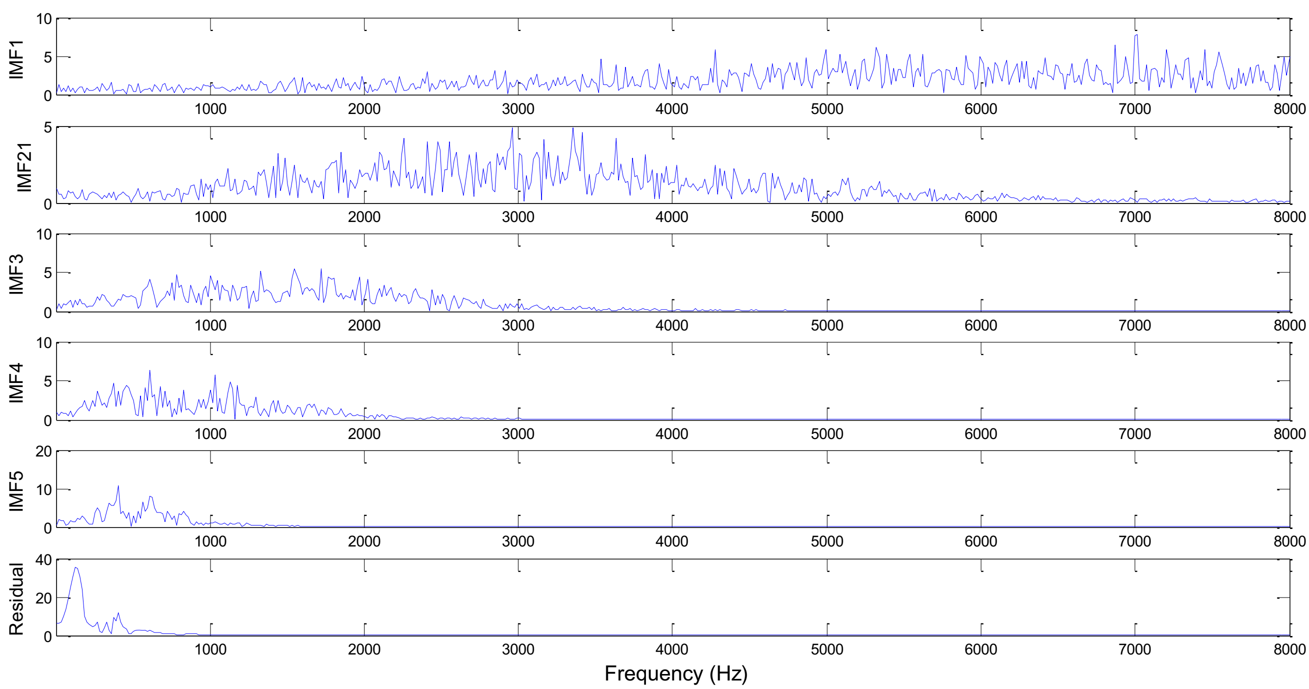
Figure 4. Procedure for efficiently extraction of the EMD-based cepstral features.

Table 2. Identification accuracy for various combination forms of the frequency.

Decomposition	Band selection	Selected bands: IMF1	IMF2	IMF3	IMF4	IMF5	Total no. of features	Accuracy
EMD	T. & E.	5:15	3: 7	4	3	1, 2 and 15	20	96.66
EMD	T. & E.	9:14	6:8	5	3 and 4	1, 2 and 15	14	95.83%
EMD	T. & E.	–	3:8	3:4	3 and 4	1, 2 and 15	12	99.16%
EMD	T. & E.	–	3: 5 and 7:8	3	3:4	1, 2 and 15	10	97.5%
CEEMD	T. & E.	–	3: 5 and 7:8	3	3:4	1, 2 and 15	10	96.66%
EMD	mRMR	–	1, 3:5, 11 and 13:14	2, 3 and 5	4	–	10	95.00%
CEEMD	mRMR	–	2, 5, 6 and 12	3:6	2:4	–	10	92.5%



(**a**) Male Speaker



(**b**) Female Speaker

Figure 5. EMD-gram of the male and the female sample frames of figure 1 in the presence of noise.

performance of the algorithm in the presence of contaminating noises.

2. Over each IMF, steps 2-4 of the original MFCC procedure are applied. But, instead of all bands, log power values are only calculated over some selectively chosen bands.
3. The selected log power values over all IMFs are concatenated in order of the corresponding IMFs and

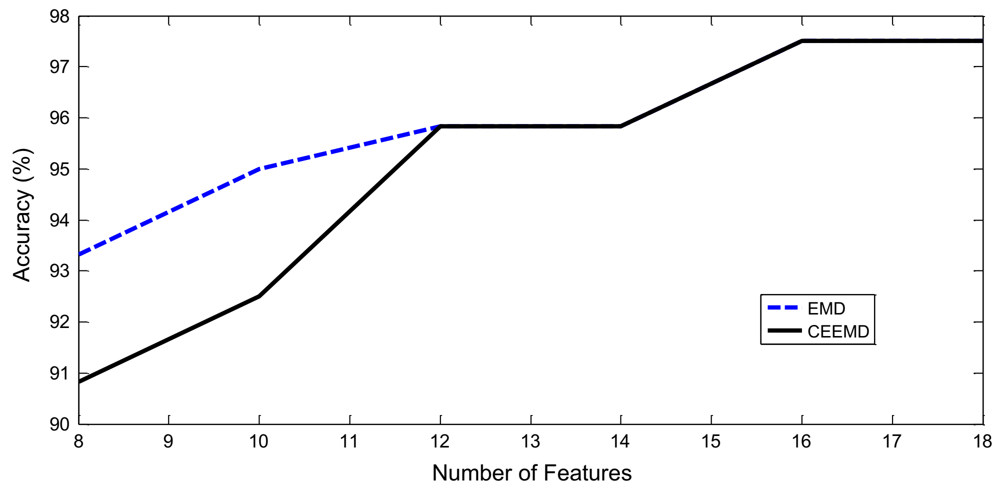


Figure 6. Effect of the number of features selected using the mRMR algorithm on the accuracy of the identification.

Table 3. Identification accuracy in the presence of Gaussian white noise.

SNR (dB)	Features Selection	MFCC (%)	EMD-MFCC T. & E. (%)	CEEMD-MFCC T. & E. (%)	EMD-MFCC mRMR (%)	CEEMD-MFCC mRMR (%)
0		50.00	50.00	64.16	55.83	61.66
3		50.00	50.00	79.16	50.00	57.50
5		51.66	50.00	80.00	50.00	50.00
7		57.50	50.00	72.50	50.00	50.00

Table 4. Identification accuracy in the presence of factory noise.

SNR (dB)	Features Selection	MFCC (%)	EMD-MFCC T. & E. (%)	CEEMD-MFCC T. & E. (%)	EMD-MFCC mRMR (%)	CEEMD-MFCC mRMR (%)
0		50.00	70.00	88.33	82.50	82.50
3		67.50	75.83	91.66	85.83	84.16
5		82.50	77.50	93.33	88.33	85.83
7		89.16	72.50	93.33	88.33	85.83

the proposed EMD-based MFCCs are subsequently extracted by applying DCT on these log power values.

To further improve the robustness of the EMD-based cepstral features, another scheme is also proposed in which a variant of the EMD, called complementary ensemble EMD (CEEMD), supersedes the EMD.

This procedure can be better followed using the block diagram of figure 4. It should be noted that in the present study frame length is set to 256 samples and a Hamming window is used for windowing. The speech signal is decomposed into 5 IMFs, over each frame. As it has been stated above, the first IMF is discarded. Each of the 4 remaining IMFs is decomposed into 15 bands, using a mel-scale triangular filter bank. On the other hand, the most efficient configuration of frequency bands as well as the total number of selected bands over all IMFs is set according to the discriminative ability of each mode and validated through some experiments, which will be reported in the next section.

4. Results

Selected sets from train and test sections of the DARPA TIMIT database [20] are used respectively for training and testing the algorithms. The selected training and testing sets contain 180 and 120 speech files, respectively. In each set, half of the speech files are uttered by female speakers and the others belong to male speakers. Support vector machine (SVM) with radial basis function (RBF) kernel with σ = 2.67 is employed for classification. For each algorithm, the classifier is first trained using all 180 training signals and then the trained model is tested over all 120 test files. The developed methods by us for cepstral feature extraction using EMD and CEEMD, with M = 10 and $\varepsilon = .002$, are compared with the original MFCC extraction procedure in both noise-free and noisy environments.

The performance of the original MFCC algorithm is highly affected by the number of frequency bands, which in turn determines the number of features. The accuracy of the gender identification algorithm using the original MFCC features, with 14 and 10 coefficients, are 99.16% and 97.5%, respectively. However, in many pattern recognition applications, identifying the most characterizing features of the data, i.e., feature selection, is critical to minimize the classification error. In step 2 of the proposed algorithm, the efficient selection of the frequency bands among 75 bands (5 IMFs, each with 15 bands) is done. For this end, we conducted some trial and error (T. & E.) experiments to study the performance of various combination forms of the frequency bands over all IMFs. Average results for some of these schemes are reported in the first five rows of table 2. As one can see, although the best result is obtained using 12 features, using 10 features leads to a better trade-off between complexity and accuracy. Furthermore, discarding the first IMF, does not degrade the performance. This is mainly due to the fact that common noises usually affect the first IMF. This phenomenon is evident from figure 5, where the EMD-gram of the male and the female sample frames of figure 1 are depicted in the presence of white Gaussian noise with 0 dB SNR. More importantly, in spite of fewer features used, utilizing the proposed EMD-based features with 12 features leads to the same result as that of the commonly used MFCC with 14 coefficients.

We also examined the minimal-redundancy-maximal-relevance (mRMR) feature selection algorithm to select 11 bands, using both EMD and CEEMD decomposition algorithms. Ten features are then obtained from the 11 selected frequency bands. Results of these tests are summarized in the last two rows of table 2. Referring to these tests, our trial and error feature selection results are in better performance than that of the mRMR algorithm. In fact, even though the mRMR algorithm shows its potential in selecting the proper bands, due to the rather thorough search we have done among various configurations, the trial and error selection of bands results in a better performance. As another experiment, the effect of changing the number of features selected using the mRMR algorithm on the accuracy of the gender identification algorithm using EMD and CEEMD-based cepstral features, is depicted in figure 6.

The effect of environmental noises on the performance of the proposed features for gender identification are studied by conducting similar tests in the presence of various noise signals. For this purpose, the accuracy of the gender identification system using 14-coefficients MFCC features are compared with that of the proposed EMD and CEEMD-based features. Averaged results, obtained in the presence of white Gaussian noise as well as factory natural noise at several signal-to-noise ratio (SNR) levels, are presented in tables 3 and 4, respectively. In these tests, the EMD and the CEEMD-based cepstral features' dimensions are set to 10, with the frequency band configurations summarized in the last four rows of table 2, while the original MFCC method uses 14 coefficients. Results obtained in the presence of some other non-Gaussian noises, e.g., pink, babble, F16 aircraft, destroyer engine and high frequency channel noises, were overall same as that of the factory noise. These results reveal the better performance of the proposed CEEMD-based methods in noisy environments, as compared to the original procedure for cepstral feature extraction.

5. Conclusion

The well-known cepstral coefficients are extracted using Fourier transform that suffers from some crucial restrictions for analyzing nonlinear and non-stationary signals such as speech. To address this problem, in this study, we investigate the application of empirical mode decomposition (EMD) in extracting more efficient and robust features for automatic gender identification. Moreover, to further alleviate the sensitivity of the proposed method to noise signals, another scheme was also developed based on a specific variant of the EMD algorithm, known as the CEEMD. In the proposed approach, the speech signal is first decomposed into its IMFs. The proposed EMD and CEEMD-based MFCCs are then extracted by applying DCT on log power values calculated over some specific bands of the IMFs, selected according to each band's discriminative ability. Since common noises are usually of high frequency, it was seen that excluding the first IMF from consideration improves the performance of the algorithm in the presence of contaminating noises. Our methods developed for cepstral feature extraction using EMD and CEEMD were compared with the original MFCC extraction procedure in both noise-free and noisy environments. It has been seen that, in spite of fewer features used, utilizing the proposed CEEMD-based features with 12 features leads to the same result as that of the commonly used MFCC with 14 coefficients. Furthermore, our proposed EMD and CEEMD-based cepstral features reveal higher robustness against environmental noises.

References

[1] Khelif K, Mombrun Y, Backfried G, Sahito F, Scarpato L, Motlicek P, Madikeri S, Kelly D, Hazzani G and Chatzigavriil E 2017 Towards a breakthrough speaker identification approach for law enforcement agencies: SIIP. In: *European Conference on Intelligence and Security Informatics*, pp. 32–39

[2] Shahin I 2017 Speaker verification in emotional talking environments based on three-stage framework. In:

International Conference on Electrical and Computing Technologies and Applications (ICECTA), pp. 1–5

[3] Zhang Y *et al* 2017 A paralinguistic approach to speaker diarisation: using age, gender, voice likability and personality traits. In: *Proceedings of the 25th ACM International Conference on Multimedia*, pp. 387–392

[4] Zeng Y *et al* 2006 Robust GMM based gender classification using pitch and RASTA-PLP parameters of speech. In: *Conference on Machine Learning and Cybernetics, International*, pp. 3376–3379

[5] Levitan S I, Mishra T and Bangalore S 2016 Automatic identification of gender from speech. In: *Proceedings* of *Speech Prosody*, pp. 84–88

[6] Harb H and Chen L 2005 Voice-based gender identification in multimedia applications. *J. Intell. Inf. Syst.* 24(2): 179–198

[7] Childers D G and Wu K 1991 Gender recognition from speech. Part II: Fine analysis. *J. Acoust. Soc. Am.* 90(4): 1841–1856

[8] Spoorthy S and Ramamurthy G 2011 Gender identification using significant intrinsic mode functions and Fourier-Bessel expansion. In: *International Conference on Signal Processing, Communication, Computing and Networking Technologies (ICSCCN)*, pp. 86–89

[9] Yücesoy E N and Nabiyev V V 2013 Gender identification of a speaker using MFCC and GMM. In: *8th International Conference on Electrical and Electronics Engineering (ELECO)*, pp. 626–629

[10] Safavi S, Russell M and Jančovič P 2018 Automatic speaker, age-group and gender identification from children's speech. *Comput. Speech Lang*. 50: 141–156

[11] Ranjan S, Liu G and Hansen J H L 2015 An i-vector PLDA based gender identification approach for severely distorted and multilingual DARPA RATS data. In: *IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*, pp. 331–337

[12] Chen O T C and Gu J 2015 Improved gender/age recognition system using arousal-selection and feature-selection schemes. In: *International Conference on Digital Signal Processing (DSP)*, pp. 148–152

[13] Huang N E *et al* 1998 The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. Lond. A: Math. Phys. Eng. Sci.* 454(1971): 903–995

[14] Flandrin P, Rilling G and Goncalves P 2004 Empirical mode decomposition as a filter bank. *IEEE Signal Process. Lett.* 11(2): 112–114

[15] Hanchuan P, Fuhui L and Ding C 2005 Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.* 27(8): 1226–1238

[16] Wu Z and Huang N E 2009 Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Adv. Adapt. Data Anal.* 1(01): 1–41

[17] Torres M E *et al* 2011 A complete ensemble empirical mode decomposition with adaptive noise. In: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4144–4147

[18] Faragallah O S 2018 Robust noise MKMFCC–SVM automatic speaker identification. *Int. J. Speech Technol.* 21(2): 185–192

[19] Khelifa M O M *et al* 2017 Constructing accurate and robust HMM/GMM models for an arabic speech recognition system. *Int. J. Speech Technol.* 20(4): 937–949

[20] TIMIT 1993 "*DARPA TIMIT-Acoustic-Phonetic Continuous Speech Corpus*," National Institute of Standards and Technology document NISTIR 4930